FULBRIGHT & JAWORSKI L.L.P.
A REGISTERED LIMITED LIABILITY PARTNERSHIP
600 CONGRESS AVENUE, SUITE 2400
AUSTIN, TEXAS 78701-3271
WWW.FULBRIGHT.COM

RHANSON@FULBRIGHT.COM
DIRECT DIAL: (512) 536-3085

TELEPHONE: (512) 474-5201
FACSIMILE: (512) 536-4598

February 24, 2003

CERTIFICATE OF MAILING
37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, DC 20231, on the date below:

February 24, 2003 — Date

Robert E. Hanson

BOX AF
COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

Re: *SN 09/602,840 entitled "METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED" by Julie A. Kirihara, et al.*
Our Ref. DEKM:180USD1; Client Ref. 51209 US 23

Commissioner:

Transmitted herewith for filing are:

1. A Brief on Appeal (an original and two copies);
2. A check for $320.00 to cover the filing fee for the appeal brief; and
3. A return postcard to acknowledge receipt of these materials. Please date stamp and mail this postcard.

If the check is inadvertently omitted, or the amount is insufficient, or should any additional fees under 37 C.F.R. §§ 1.16 to 1.21 be required for any reason relating to the enclosed materials, or should an overpayment be included herein, the Commissioner is authorized to deduct or credit said fees from or to Fulbright & Jaworski L.L.P. Account No.: 50-1212/DEKM:180USD1.

Respectfully submitted,

03/05/2003 MBIZUNES 00000143 09602840
01 FC:1402 320.00 OP

Robert E. Hanson
Reg. No. 42,628

REH/vv
Enclosures
25266321.1

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:
Julie A. Kirihara, et al.

Serial No.: 09/602,840

Filed: June 23, 2000

For: METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED

Group Art Unit: 1638

Examiner: Stuart F. Baum

Atty. Dkt. No.: DEKM.180USD1

CERTIFICATE OF MAILING
37 C.F.R. §1.8

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as First Class Mail in an envelope addressed to: Commissioner for Patents, Washington, D.C. 20231, on the date below:

February 24, 2003
Date

Robert E. Hanson

BRIEF ON APPEAL

TABLE OF CONTENTS

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Julie A. Kirihara, et al.	Group Art Unit: 1638
Serial No.: 09/602,840	Examiner: Stuart F. Baum
Filed: June 23, 2000	Atty. Dkt. No.: DEKM:180USD1
For: METHOD FOR ALTERING THE NUTRITIONAL CONTENT OF PLANT SEED	

BRIEF ON APPEAL

BOX AF
Commissioner of Patents
Washington, D.C. 20231

Sir:

Appellants hereby submit an original and two copies of this Appeal Brief in response to the Final Office Action dated September 27, 2002. The fee for filing this Appeal Brief is attached hereto. This Brief is filed pursuant to the Notice of Appeal mailed December 23, 2002. The date for filing the instant Brief is February 28, 2003, based on the receipt of the Notice of Appeal by the Patent and Trademark Office on December 30, 2002.

No additional fees are believed due in connection with the instant paper. However, should any other fees be due, or the attached fee be deficient or absent, the Commissioner is authorized to withdraw the appropriate fee from Fulbright & Jaworski L.L.P. Deposit Account No. 50-1212/DEKM:180USD1. Please date stamp and return the enclosed postcard to evidence receipt of this document.

I. REAL PARTIES IN INTEREST

The real party in interest is Monsanto Company, the parent of wholly-owned subsidiary DeKalb Genetics Corporation, the assignee of this application.

II. RELATED APPEALS AND INTERFERENCES

There are no interferences or appeals for related cases.

III. STATUS OF THE CLAIMS

Claims 1-100 were filed with the original application on June 23, 2000. Claims 1-71, 74-77, 80-83, 85, 97 and 92-93 were canceled and claims 78, 84, 86, 88-91, 94, 95, and 100 were amended in a Preliminary Amendment filed concurrently with the application. Claims 100-114 were added in a Supplemental Preliminary Amendment, mailed in the case on November 1, 2001.

Claims 72, 73, 78, 79, 84, 86, 88-91 and 94-110 were indicated in the first Office Action as pending in the case. It is thus believed by Appellants that the Examiner had renumbered claims 100-114 to 96-110, due to failure to use claim numbers 96-99 in the application as filed and in subsequent Preliminary Amendments. Added claims 100-114 have thus been renumbered to 96-110.

Claims 72, 73, 78, 79, 97 and 98 were amended in Appellants response to the first Office Action, mailed June 13, 2002. Claims 72-73, 78-79, 84, 86, 88-91, 94-110 were pending at the time of the final Office Action mailed in the case on September 27, 2002 and were finally rejected. No amendments have been filed after the final Office Action. The final rejection of claims 72-73, 78-79, 84, 86, 88-91, 94-110 is the subject of the instant appeal. A copy of the appealed claims is attached hereto as APPENDIX 1.

IV. STATUS OF AMENDMENTS

No amendments were made subsequent to the final Office Action.

V. SUMMARY OF THE INVENTION

The invention relates to transgenic maize plants having increased starch content and extractability and methods of use thereof. Specification at page 6, line 13 to page 7, line 2. Increased starch content and extractability is obtained by the expression of RNA molecules that are substantially identical or complementary to 19kD or 22kD α-zein plant seed storage proteins, thereby decreasing the amount of corresponding seed storage protein and increase in starch content and extractability in the cells of the plant. Specification at page 6, line 13 to page 7, line 23.

VI. ISSUES ON APPEAL

(1) Are claims 72-73, 78-79, 84, 86, 88-91 and 94-110 properly rejected under 35 U.S.C. §112, first paragraph, as not being enabled by the specification?

(2) Are claims 88-89, 90, 91 and 102-107 properly rejected under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out the subject matter which Applicant regards as the invention?

VII. GROUPING OF THE CLAIMS

The claims stand or fall together.

VIII. SUMMARY OF THE ARGUMENT

A. The Examiner has failed to establish a *prima facie* case of lack of enablement under 35 U.S.C. § 112, first paragraph, and improperly ignores evidence submitted by Appellant affirmatively demonstrating the enablement of the claims.

B. The terms "substantially identical" and "substantially complementary" are not indefinite under 35 U.S.C. § 112, second paragraph, as they are used in the claims as one of skill in the art would readily understand the meaning of the terms when viewed in light of the language of the claim and teaching of the specification.

VIII. ARGUMENT

A. Rejection Under 35 U.S.C. §112, First Paragraph

Claims 72-73, 78-79, 84, 86, 88-91 and 94-110 were finally rejected under the first paragraph of 35 U.S.C. §112 as allegedly not being enabled by the specification. In particular, the Examiner alleged that the specification does not teach one of skill in the art how to increase the starch content or starch extractability or kernel hardness of seeds. It is stated that the specification only teaches how to make maize seeds with decreased amounts of the amino acid leucine and increased lysine by transforming *Zea mays* plants with SEQ ID NOs:1 and 2 operably linked to a Z10 promoter.

1. The Examiner Has Failed to Establish a Prima Facie Case of Lack of Enablement

The claims are fully enabled by the specification. No evidence has been put forth by the Examiner sufficient to doubt enablement. In this regard, Appellants note that the Examiner has apparently placed the burden on Applicant to affirmatively establish enablement. For example, on page 3 of the final Office Action, the Examiner states that "Applicant does not address the issue of enablement, rather Applicant only addresses the validity or applicability of references

cited by the Examiner in support of undue experimentation to make and/or use the claimed invention." However, what Appellants had done in the previous response to the first Office Action is demonstrate that the Examiner had failed to establish any basis to doubt enablement, while affirmatively demonstrating why the very references cited by the Examiner supported enablement. The examiner has, therefore, failed to meet the burden of establishing a *prima facie* case of lack of enablement. An unsupported allegation of lack of enablement does not meet the PTO's burden under 35 U.S.C. § 112, first paragraph, "[o]therwise, there would be no need for the applicant to go to the trouble and expense of supporting his presumptively accurate disclosure." *In re Marzocchi*, 169 U.S.P.Q. 367, 369-70 (CCPA 1971).

2. *The References Cited by the Examiner Support the Enablement of Appellants Claims*

In Appellants response to the first Office Action, the lack of any support for the enablement rejection was demonstrated. The Examiner attempted to support the rejection by citing two references: Coleman *et al.* (1997) (EXHIBIT A) and Marks *et al.* (1985) (EXHIBIT B). Coleman *et al.* was cited as showing that high-lysine mutants exhibiting a reduction of α-zein content were "concomitant with an inferior endosperm quality." It was thus suggested that, because of the supposedly inferior endosperm quality, "reducing the α-zein protein content of maize seeds using the strategy of Applicants, will not increase the starch content or starch extractability of maize seeds." However, the reference does the very opposite and demonstrates the enablement of the claims. This "inferior" endosperm is in fact a ***soft and starchy endosperm***. (See EXHIBIT A, p. 7094, paragraph 2). The reference therefore shows the direct correlation between increased lysine, decreased α-zein and soft and starchy endosperm. The first Office

Action acknowledges this on page 5.[1] Demonstration of this correlation supports enablement, given that the final Office Action has already acknowledged that Appellants have demonstrated enablement for increasing lysine. The fact that Coleman *et al.* also demonstrated successful expression of a 24 kDa α-zein gene to induce the mutant phenotype provides still further evidence of enablement of the instant claims.

It is further noted that whether the endosperm is subjectively "inferior" or not is irrelevant to enablement. What is considered inferior for a plant used for one purpose, for example, for human consumption, is not necessarily the same as for a plant used for production of corn starch. Further, an "inferior" endosperm does not equate to an inability to increase starch content or extractability, as described in Coleman *et al.*

The other reference cited in the rejection, Marks *et al.* (1985) (EXHIBIT B), was cited as allegedly indicating that there are many different forms of 19 kD and 22 kD α-zeins with divergent and unpredictable functions. However, this reference also ***supports*** the enablement of the invention. Marks *et al.* states in the abstract that:

> A comparison of the DNA and protein sequences of a group of zein cDNA clones reveals that they ***share extensive sequence homology and probably originated from a common ancestral gene***. A comparison of clones corresponding to Mr 22,000 polypeptides shows that they are 92% homologous, while five clones corresponding to Mr 19,000 zein vary in homology from 75 to 95%. (emphasis added)

Marks *et al.*, therefore, demonstrates the common structural characteristics shared among 19 kD and 22 kD α-zeins. In the first paragraph of the Discussion section of Marks *et al.*, it is indicated that cDNA sequences among the 19 kD and 22 kD group of α-zein sequences are 75 to 95% and 92% homologous, respectively. Further, Marks *et al.* provides sequence information and

---

[1] See middle paragraph: "Two 'high –lysine' mutants were identified, opaque2 (o2) and floury2 (fl2) which have a ***higher lysine content due to a reduction in the α-zein protein content*** of the endosperm."

comparisons among 19 kD and 22 kD α-zeins. The disclosure of Marks *et al.* thereby demonstrates the shared structural characteristics of the 19 kD and 22 kD α-zein seed storage proteins.

No basis has been provided by the Examiner to indicate that different isoforms have "divergent functions." The high degree of homology and indicated common ancestor among zeins strongly contradicts this assertion. Additionally, there is no support in the Marks *et al.* reference for the contention that any of the mRNA isoforms described encode proteins other than zeins, the function of which is that of a seed storage protein. The reference, therefore, does ***not*** indicate divergent functions, but rather indicates the shared original and conserved structure of the 19 kD and 22 kD α-zein plant seed storage proteins.

3. The Examiner Improperly Ignored Data Submitted by Appellants Affirmatively Demonstrating Enablement of the Claims

In doubting Appellants evidence of enablement, the Examiner noted that, of four independent lines representing 8 samples in Table V on page 78 of the specification, only three samples exhibited a statistical increase in lysine (*e.g.*, show expression of the zein seed storage protein) compared to a control. However, this is irrelevant to enablement. What is relevant is that the Application teaches one of skill in the art to make and use the claimed invention without undue experimentation. An indicated success rate of 3/8 cannot reasonably indicate undue experimentation, as it would be well within the capabilities of one of skill in the art to readily produce many lines expressing 19 kD and 22 kD α-zein plant seed storage proteins without undue experimentation using the methodology described in the specification. No basis has been provided to suggest why one of skill in the art would reasonably conclude otherwise. The fact

that three of eight samples exhibited a statistical increase in lysine demonstrates the repeatability and enablement of the technique.

Further evidence of enablement of the claims is provided at page 83 of the specification. There it is described that endosperm cells in a maize kernel are made up primarily of large starch granules and protein sequestered in protein bodies. It is shown that a reduction in the number of protein bodies in endosperm cells derived from a transformant produced using a zein antisense construct was achieved. This was shown by light microscopy and the results given in FIG. 9. As can be seen in the figure, the results demonstrated a decrease in the amount of seed storage protein and therefore an increase in the relative starch content of the kernel.

In conclusion, the Examiner has failed to set forth any basis for doubting the enablement of the claims. Appellants have further affirmatively set forth evidence of the enablement of the claims that has not been contradicted by the Examiner. The rejection must therefore be reversed.

B. Rejection of Claims Under 35 U.S.C. §112, Second Paragraph

Claims 88-89, 90, 91 and 102-107 were finally rejected under 35 U.S.C. §112, second paragraph, as allegedly being indefinite for failing to particularly point out the subject matter which Applicant regards as the invention. In particular, the Action rejects the claims as being indefinite for the recitation of "substantially complementary to all or a portion" and "substantially identical to all or a portion." In particular, it is stated that the recited phrases are vague and unclear and do not specify what portion or percent of the sequence applicants are referring to.

In response, it is noted that the terms "substantially identical" and "substantially complementary" are defined in the specification at page 12, lines 11-24. The use of the terms in the claims is thus not indefinite. Still further, the claims define the terms by indicating that expression of the seed storage protein RNA decreases the amount of seed storage protein. Thus the meaning of the terms to one of skill in the art is clear, *e.g.*, that substantially identical to or complementary to all or a portion refers to those sequences that are complementary such as to hybridize with the seed storage protein mRNA *in vivo* to cause antisense suppression. Such a sequence could not represent a single homologous base pair, as it is well known to those of skill in the art that in order to form the type of stable complex required for antisense suppression, longer stretches of complementary sequences are required.

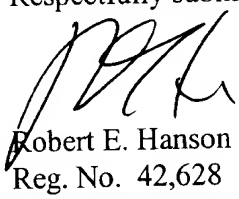
The cited terms are fully definite in compliance with the statute. The test for definiteness under 35 U.S.C. § 112, second paragraph is whether "those skilled in the art would understand what is claimed when the claim is read in light of the specification." *Orthokinetics, Inc. v. Safety Travel Chairs, Inc.*, 806 F.2d 1565, 1576, 1 USPQ2d 1081, 1088 (Fed. Cir. 1986). If one skilled in the art is able to ascertain the meaning of the claim, 35 U.S.C. § 112, second paragraph, is satisfied. *Id.* Given the definition in the specification and claim language, the referenced terms fully meet this requirement.

In conclusion, Appellants note that the rejected claims are fully definite in compliance with 35 U.S.C. § 112, second paragraph. The rejection must therefore be reversed.

IX. CONCLUSION

It is respectfully submitted, in light of the above, that none of the appealed claims lack enablement or are indefinite. Therefore, Appellants respectfully request that the Board reverse the pending grounds for rejection.

Respectfully submitted,



Robert E. Hanson
Reg. No. 42,628
Attorney for Appellants

FULBRIGHT & JAWORSKI L.L.P
600 Congress Avenue, Suite 2400
Austin, Texas 78701
(512) 536-3085

Date: February 24, 2003

APPENDIX 1: APPEALED CLAIMS

72. A fertile transgenic *Zea mays* plant having an increased starch content, the genome of which is stably augmented by a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the transgenic plant in an amount sufficient to decrease the amount of said seed storage protein and increase starch content in the cells of a plant which only differ from the cells of said transgenic plant in that said preselected DNA sequence is absent, and wherein said preselected DNA sequence is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation.

73. A fertile transgenic *Zea mays* plant, the seeds of which have an increased starch extractability, the genome of said plant which is stably augmented by a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein plant seed storage protein, wherein the preselected DNA sequence is expressed in the seeds of the transgenic plant in an amount sufficient to decrease the amount of said seed storage protein and increase the starch extractability of the seed relative to the amount of said seed storage protein and starch extractability in the seeds of a plant which only differ from the seeds of said transgenic plant in that said preselected DNA sequence is absent, and wherein said preselected DNA sequence is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation.

78. A seed derived from the plant of claim 72 or 73, wherein the seed comprises said preselected DNA sequence.

79. A progeny plant derived from the seed of claim 78, wherein the plant comprises said preselected DNA sequence.

84. The transgenic plant of claim 72 or 73, wherein the promoter comprises the 10 kD zein promoter.

86. The transgenic plant of claim 72 or 73, wherein the promoter comprises the 27kD zein promoter.

88. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding 19 kD α-zein protein.

89. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 22 kD α-zein protein.

90. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule, substantially identical to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 19 kD α-zein protein.

91. The transgenic plant of claim 72 or 73, wherein the preselected DNA sequence, which encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a seed storage protein, encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 22 kD α-zein protein.

95. The transgenic plant of claim 72 or 73, wherein the cell is transformed by a method selected from the group consisting of electroporation, microinjection, microprojectile bombardment, and liposomal encapsulation.

96. The transgenic plant of claim 78 or 79, further comprising stably transforming the cells with at least one selectable marker gene.

97. A method of producing a *Zea mays* seed with an increased starch content, comprising:

(a) growing a transgenic *Zea mays* plant, the genome of which is augmented with a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary to an mRNA encoding a 19kD or a 22kD α-zein seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the *Zea mays* plant in an amount sufficient to decrease the amount of seed storage protein; and

(b) selecting a seed from the transgenic *Zea mays* plant, wherein the seed has an increased amount of starch relative to the amount of starch in a seed selected from a plant which does not comprise said preselected DNA sequence.

98. A method of obtaining starch from a *Zea mays* seed, comprising:

(a) growing a transgenic *Zea mays* plant, the genome of which is augmented with a preselected DNA sequence encoding an RNA molecule which is substantially identical, or complementary, to an mRNA encoding a 19kD or a 22kD α-zein seed storage protein, wherein the preselected DNA sequence is expressed in the cells of the *Zea mays* plant in an amount sufficient to decrease the amount of seed storage protein;

(b) obtaining seed from said plaint; and

(c) extracting starch from the seed.

99. The method of claim 97 or 98 wherein the preselected DNA sequence is operably linked to a promoter functional in plant cells.

100. The method of claim 99 wherein the promoter comprises the 10 kD zein promoter.

101. The method of claim 99 wherein the promoter comprises the 27 kD zein promoter.

102. The method of claim 97 or 98 wherein the preselected DNA sequence encodes an RNA molecule which is substantially identical to all or a portion of the mRNA encoding a seed storage protein.

103. The method of claim 97 or 98 wherein the preselected DNA sequence encodes an RNA molecule which is substantially complementary to all or a portion of the mRNA encoding a seed storage protein.

104. The method of claim 102 wherein the preselected DNA sequence encodes an RNA molecule substantially identical to all or a portion of mRNA encoding a 19 kD α-zein protein.

105. The method of claim 102 wherein the preselected DNA sequence encodes an RNA molecule substantially identical to all or a portion of an mRNA encoding a 22 kD α-zein protein.

106. The method of claim 103 wherein the preselected DNA sequence encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 19 kD α-zein protein.

107. The method of claim 103 wherein the preselected DNA sequence encodes an RNA molecule substantially complementary to all or a portion of an mRNA encoding a 22 kD α-zein protein.

108. The method of claim 97 or 98 wherein the genome of the transgenic *Zea mays* plant is further augmented with a DNA sequence encoding a polypeptide that provides the transgenic *Zea mays* plant with increased kernel hardness.

109. The method of claim 97 or 98 wherein the transgenic *Zea mays* plant is produced from cells transformed by a method selected from the group consisting of electroporation, microinjection, microprojectile bombardment, and liposomal encapsulation.

110. The method of claim 97 or 98 wherein the genome of the transgenic *Zea mays* plant is further augmented with at least one selectable marker gene.

APPENDIX 2: EXHIBITS

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 7094–7097, June 1997
Plant Biology

Expression of a mutant α-zein creates the *floury2* phenotype in transgenic maize

(endosperm/prolamin)

CRAIG E. COLEMAN*†, AMY M. CLORE*, JERRY P. RANCH‡, REGINA HIGGINS‡, MAURICIO A. LOPES§, AND BRIAN A. LARKINS*¶

*Department of Plant Sciences, University of Arizona, Tucson, AZ 85721; ‡Pioneer Hi-Bred International, Inc., 7300 NW 62nd Avenue, Johnston, IA 50131-1004; and §Centro Nacional de Pesquisa de Milhoe Sorgo–Empresa Brasieira de Pesquisa Agropecuaria, Caixa Postal 151, 35700 Sete Lagoas, MG, Brazil

Contributed by Brian A. Larkins, May 6, 1997

ABSTRACT **The maize *floury2* mutation results in the formation of a soft, starchy endosperm with a reduced amount of prolamin (zein) proteins and twice the lysine content of the wild type. The mutation is semidominant and is associated with small, irregularly shaped protein bodies, elevated levels of a 70-kDa chaperone in the endoplasmic reticulum, and a novel 24-kDa polypeptide in the zein fraction. The 24-kDa polypeptide is a precursor of a 22-kDa α-zein protein that is not properly processed. The defect is due to an alanine-to-valine substitution at the C-terminal position of the signal peptide, which causes the protein to be anchored to the endoplasmic reticulum. We postulated that the phenotype associated with the *floury2* mutation is caused by the accumulation of the 24-kDa α-zein protein. To test this hypothesis, we created transgenic maize plants that produce the mutant protein. We found that endosperm in seeds of these plants manifests the *floury2* phenotype, thereby confirming that the mutant α-zein is the molecular basis of this mutation.**

Zeins are prolamin storage proteins that accumulate in the endosperm of maize (*Zea mays* L.) seeds. They are composed of four different types of polypeptides, classified as α-, β-, γ-, and δ-zeins (1). Accretions of zein proteins form spherical protein bodies within the lumen of the endoplasmic reticulum (ER), and there is a distinct spatial arrangement of these proteins within a protein body: β- and γ-zeins are located on the periphery, whereas α- and δ-zeins are found in the interior (2, 3). Collectively, the zein proteins are rich in glutamine and proline, but they lack lysine and tryptophan. Because zeins constitute such a large proportion of the total seed protein (60–70%), the amino acid composition of these proteins causes the grain to be of inferior nutritional quality for monogastric animals.

Efforts to improve the protein quality of maize seed have focused on mutants in which zein synthesis is reduced and the lysine content is increased. The first "high-lysine" mutants to be identified were opaque2 (*o2*) and floury2 (*fl2*) (4, 5). Unfortunately, the favorable nutritional quality of these mutants is offset by the inferior physical properties of their soft, starchy endosperms. It appears that the starchy endosperm of the *o2* and *fl2* mutants is caused by changes in the nature of their protein bodies. The *o2* mutation affects a transcriptional activator of a subset of α-zein genes, leading to a reduction in α-zein protein synthesis and the formation of protein bodies that are significantly smaller than normal (6–9). The *fl2* mutation, which is semidominant, causes a decrease in synthesis of all classes of zeins, and the resultant protein bodies are not only smaller than normal, but they are also asymmetrical and misshapen (10, 11). Another feature of *fl2* endosperm is the overexpression of the ER-resident binding protein (BiP), which becomes deposited at the periphery of the mutant protein bodies (12–15).

We have postulated that the phenotype associated with *fl2* is caused by the accumulation of a novel 24-kDa α-zein protein (16, 17). This hypothesis is partially based on tight linkage between the gene encoding the 24-kDa protein and the *fl2* locus, but it is also consistent with the abnormal structure of the protein. The mutant protein is 2 kDa larger than expected, because of a defect in its processing following targeting to the ER lumen. An alanine-to-valine substitution at the C-terminal position of the signal peptide prevents its removal, thereby anchoring the protein to the lumenal face of the ER membrane (18). To investigate whether this protein is responsible for the mutant phenotype, we transformed normal maize plants with the gene encoding the mutant 24-kDa α-zein protein. We show here that seeds of these plants manifest the key phenotypic characteristics associated with the *fl2* mutation.

MATERIALS AND METHODS

Transformation of Maize Plants. Maize embryos were cotransformed with the 24-kDa α-zein gene in plasmid pCC515 (17) and the bacterial bialaphos (*BAR*) gene (19) as a selectable marker. The 24-kDa α-zein gene in pCC515 is flanked by 3.0 kb of 5′ and 3.7 kb of 3′ noncoding sequences. The selectable marker gene plasmid is of the form *ubi::ubiintron::BAR::pinII*, where *ubi* is a maize ubiquitin promoter and first intron and *pinII* is the potato protease inhibitor II 3′ noncoding sequence. Plasmid DNAs were delivered by microprojectile bombardment to embryogenically responsive, immature embryos from the maize variety High Type II (20). Embryos were recovered from herbicide-resistant calli and grown to maturity. Transgenic plants were outcrossed as the female parent to an inbred line, and the progeny were scored for the floury kernel phenotype.

PCR Amplification of the 24-kDa α-Zein Gene. Genomic DNA was extracted from seedlings germinated from normal and floury kernels (21). The DNA was amplified by PCR using primers corresponding to the signal peptide coding sequences of the 24-kDa α-zein gene (5′-GCCCTTTTAGTGAGCG-CAACAAATGTG-3′) and coding sequences for the seventh α-helical repeat of the protein (5′-GCAGGGTTTGCCAT-AGCTAGCTGATG-3′). Products were separated by electrophoresis in 1% agarose gels, stained with ethidium bromide, and photographed with a Polaroid DS-34 camera.

Protein Extraction from Maize Flour and Immunoblotting. A portion of each endosperm was cut from the seed prior to

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BiP, binding protein; ER, endoplasmic reticulum.

†Present address: Department of Botany, Brigham Young University, Provo, UT 84602.
¶To whom reprint requests should be addressed.

germination and converted into a fine flour using a ball mill. Proteins were extracted from the meal and separated according to their solubility in 70% alcohol (22). Alcohol-soluble (zein) proteins were separated by SDS/PAGE, blotted onto nitrocellulose, and immunoreacted with rabbit anti-α-zein polyclonal antibody (23). Alcohol-insoluble proteins were separated by SDS/PAGE, blotted, and immunoreacted with a rabbit anti-BIP polyclonal antibody (12). Goat anti-rabbit alkaline phosphatase conjugate was used for indirect detection of α-zein and BiP on the immunoblots (24).

Fixation and Embedding of Endosperms and Electron Microscopy. Seeds were harvested 18 days after pollination from a self-fertilized maize plant that was hemizygous for the 24-kDa α-zein transgene. Endosperms were fixed, embedded, sectioned, and viewed with a transmission electron microscope as described elsewhere (11).

RESULTS

Detection of the Transgene in Transformed Maize Seedlings. Transgenic maize plants were generated by a biolistic method using microprojectiles coated with plasmid pCC515 (17), which contains the 24-kDa α-zein gene within a genomic DNA fragment, and a plasmid containing the *BAR* gene (20), which confers resistance to the herbicide Basta. Plants from 25 herbicide-resistant events that were recovered from the transformation were crossed as females to an untransformed inbred line. F_1 progeny from 17 of these crosses segregated approximately 1 to 1 for floury-appearing kernels, consistent with the presence of a single site of transgene integration. To determine whether this phenotype was associated with the insertion of the 24-kDa α-zein gene, we used PCR primers for the coding region of the 24-kDa α-zein gene to amplify genomic DNA from F_1 seedlings. A 560-bp fragment was produced from DNA of the floury seedlings (Fig. 1*A*, lanes 4, 6, and 8), but not the wild-type seedlings (Fig. 1*A*, lanes 3, 5, and 7). A fragment of similar size was amplified from W64A*fl2* DNA (Fig. 1*A*, lane 9), but not from DNA of untransformed seedlings (Fig. 1*A*, lanes 1 and 2).

Immunodetection of α-Zein Proteins from Seeds of Transgenic Plants. An immunoblot of α-zein proteins from mature seeds was prepared to determine whether insertion of the 24-kDa α-zein gene resulted in synthesis of this protein (Fig. 1*B*). The blot shows that the appearance of a 24-kDa protein band, indicated by arrowheads in lanes 4, 6, and 8 of Fig. 1*B*, was always associated with the floury phenotype and insertion of the transgene. The 24-kDa protein band in these samples is similar to a band of identical molecular mass that was found in W64A*fl2* zein (indicated by an arrowhead in lane 9 of Fig. 1*B*). The 24-kDa α-zein was not present in samples from normal progeny or seeds of untransformed plants (Fig. 1*B*, lanes 1, 2, 3, 5, and 7). Three bands with molecular masses greater than 24 kDa were detected in all of the samples. In W64A*fl2*, these polypeptides are glycosylated forms of 19-kDa α-zein proteins that are products of a gene(s) closely linked to the *fl2* locus (J. W. Gillikin and R. S. Boston, personal communication). Apparently, the same glycoproteins are present in the maize line used for the transformation, although they had no effect on the kernel phenotype.

Analysis of BiP Expression in Seeds of Transgenic Plants. One phenotypic characteristic associated with the *fl2* mutation is overexpression of the 70-kDa ER-resident chaperone, BiP. Using anti-BiP antibody, an immunoblot of the alcohol-insoluble proteins showed that the amount of BiP in mature

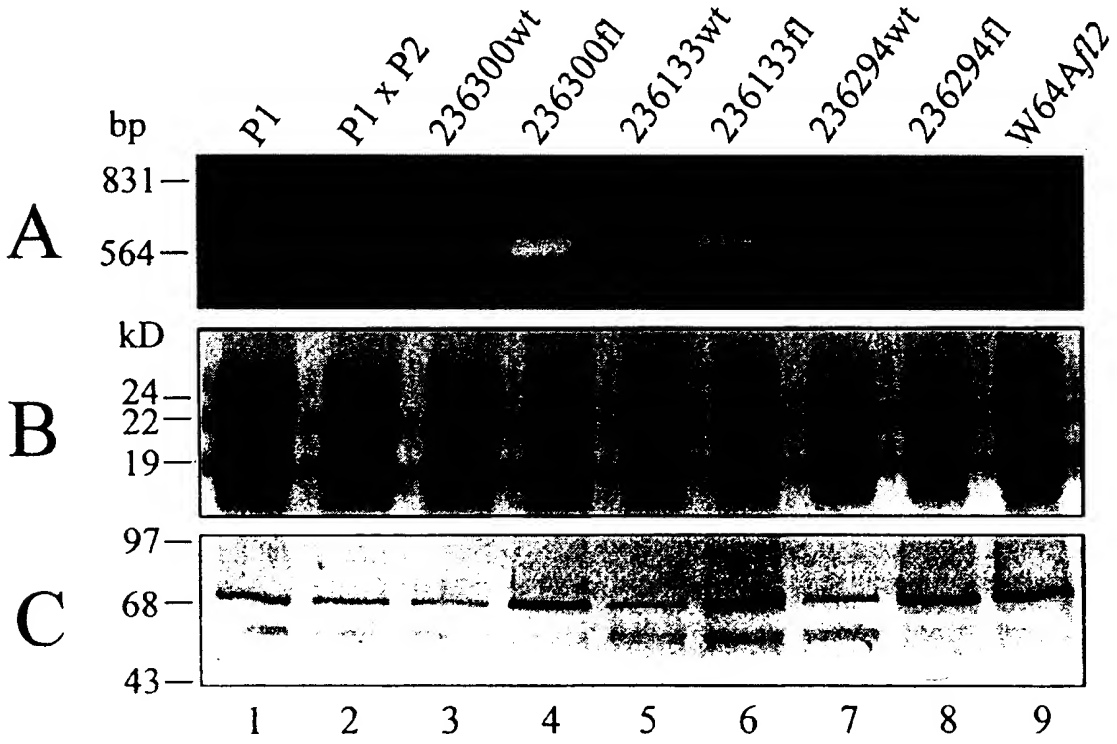


FIG. 1. Transformation of the 24-kDa α-zein gene into transgenic maize plants leads to synthesis of the encoded protein and overexpression of BiP. Analysis of 3 of 17 independent transgenic lines is shown. One normal and one floury kernel from transgenic lines 236300 (lanes 3 and 4), 236133 (lanes 5 and 6), and 236294 (lanes 7 and 8) were analyzed and compared with kernels from the untransformed (lane 1, P1), the untransformed parent outcrossed to a normal inbred (lane 2, P1 × P2), and W64A*fl2* (lane 9) plants. (*A*) Using PCR primers specific to the 24-kDa α-zein gene, a 560-bp DNA fragment was amplified from genomic DNA of seedlings germinated from floury kernels, but not wild-type kernels. (*B*) An immunoblot of α-zeins shows a 24-kDa protein band (indicated by arrowheads) in samples from floury kernels, but not wild-type kernels. (*C*) An immunoblot of the 70-kDa maize homolog of BiP.

transgenic kernels was comparable to that of W64A*fl2* (Fig. 1*C*, lanes 4, 6, 8, and 9). Normal quantities of BiP were detected in wild-type siblings and untransformed kernels (Fig. 1*C*, lanes 1, 2, 3, 5 and 7).

Protein Body Structure in the Endosperms of Seeds of Transgenic Plants. Another phenotypic characteristic associated with the *fl2* mutation is the formation of misshapen protein bodies. Protein bodies in the developing endosperm of seeds not expressing the 24-kDa α-zein gene are circular in cross-section and are relatively discreet; similar to the wild type (Fig. 2*A*; ref. 2), whereas those from seeds expressing the gene have a convoluted shape and aggregate into large masses of protein, similar to W64A*fl2* (Fig. 2*B*; ref. 11). Analysis at higher magnification revealed that protein bodies in the developing endosperm of seeds not expressing the 24-kDa α-zein gene contain darkly staining zein proteins primarily on the periphery of the protein bodies, as is seen in wild type (Fig. 2*C*, arrowheads; ref. 2). In contrast, those from endosperms expressing the gene have many regions of darkly staining zein in their interior (Fig. 2*D*, arrowheads; ref. 11), indicating an abnormal organization of α-, β-, and γ-zeins within the protein body.

DISCUSSION

Previous studies provided strong genetic evidence implicating a mutant 24-kDa α-zein protein as the cause of the phenotype

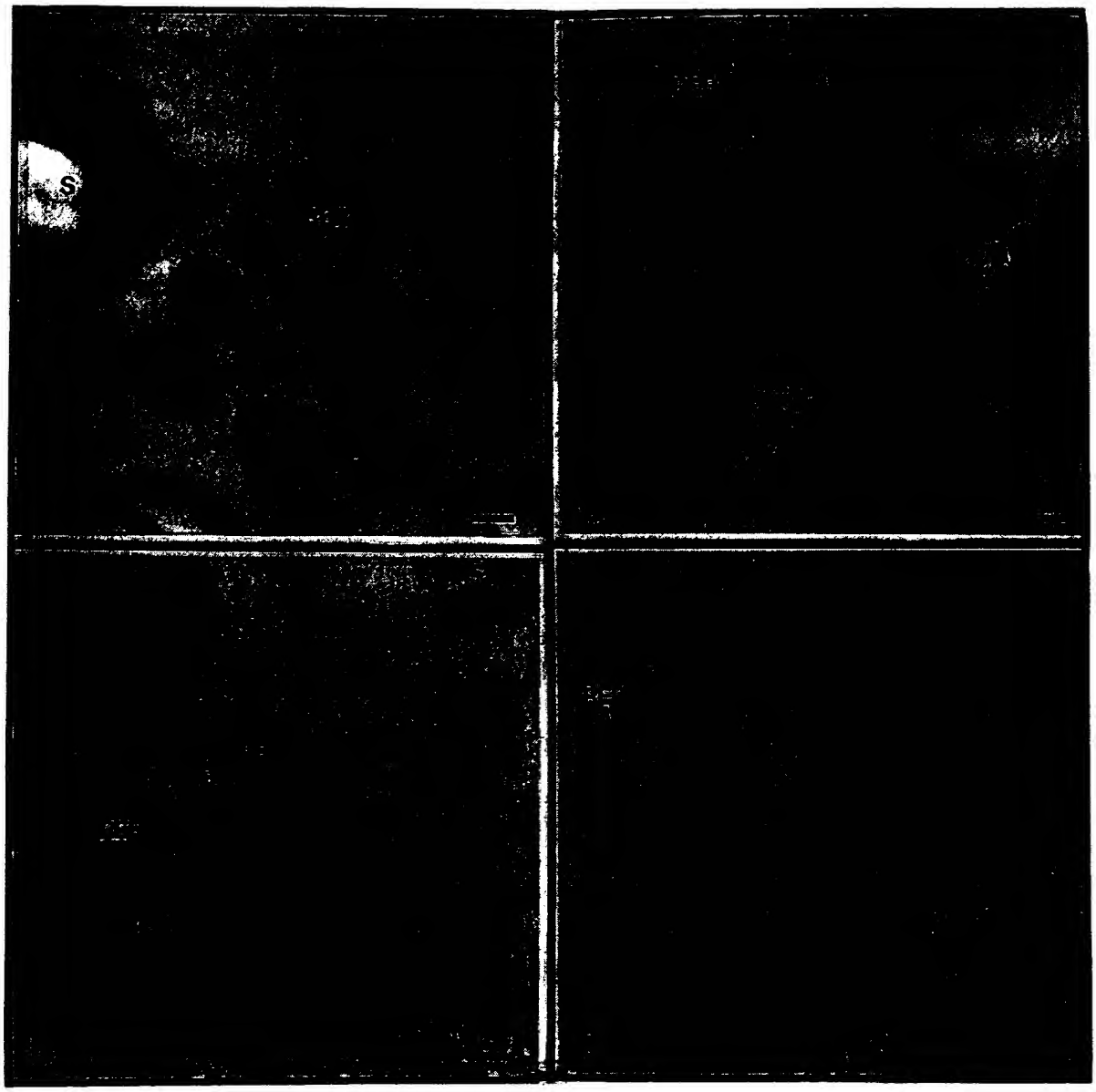


FIG. 2. Comparison of protein bodies from wild-type and transgenic maize endosperm. Protein bodies in the endosperm of wild-type seeds are round (*A* and *C*) compared with the misshapen protein bodies in the endosperm of seeds expressing the 24-kDa α-zein gene (*B* and *D*). Comparison of these protein bodies at lower magnification revealed that they appear as discreet spheres (circles in cross-section) in cells not expressing the 24-kDa α-zein gene (*A*), whereas they have a convoluted shape and are aggregated in cells expressing this gene (*B*). Comparison of these protein bodies at higher magnification highlights their differences in the internal structure. In cells not expressing the gene, the darkly staining zein proteins are found primarily on the periphery of the protein bodies (*C*, arrowheads), whereas in cells expressing the gene, the protein bodies have many locules of darkly staining zein in their interior (*D*, arrowheads). The irregularities seen in *B* and *D* are similar to those found in protein bodies of W64A*fl2*. PB, protein body; RER, rough endoplasmic reticulum; S, starch grain. Bars = 0.5 μm.

associated with the maize *fl2* mutation (16, 17). The experiments described here show that expression of the mutant gene in transgenic maize plants leads to the accumulation of the 24-kDa α-zein protein, which is correlated to the overexpression of BiP and the appearance of malformed protein bodies. Because these characteristics of the seed from the transgenic plants resemble the phenotype of the *fl2* mutant, our observation conclusively demonstrates that *fl2* is a structural mutation in a 22-kDa α-zein gene.

The mechanism whereby the 24-kDa α-zein protein disrupts the normal development of the protein body, leading to an altered texture of the mature endosperm, is not known. One possible explanation is that the mutant hydrophobic α-zein protein is anchored to the ER membrane and remains on the outer surface of the protein body, thereby disrupting protein interactions that are important to the maintenance of the spherical shape. Normally, α-zein proteins become sequestered in the interior of the protein body (2). This hypothesis is supported by evidence showing that the 24-kDa α-zein protein is associated with the membrane fraction following translation and translocation in the presence of microsomes (18).

Our results provide evidence of the utility of maize transformation for the analysis of genetic mechanisms. The amount of the 24-kDa α-zein protein detected in the transgenic floury kernels and the W64A*fl2* kernel was similar, suggesting that the expression of the transgene may be comparable to that of the native gene. This result is encouraging, because it demonstrates that transgene expression in tissues of transformed maize plants can be adequately controlled by native promoters. Furthermore, the 3 kb of 5′ and 3.7 kb of 3′ noncoding sequence included in the α-zein genomic clone must be sufficient for directing appropriate temporal and spatial expression in the endosperm; however, it is not known whether the gene is transcribed in other tissues of the transgenic plants.

Traditionally, techniques for stable transfer of DNA to monocotyledonous cereals have lagged behind similar work with dicotyledonous plants (25). Transformation of dicot species usually involves gene delivery through infection with *Agrobacterium tumefaciens*, but monocots are not readily amenable to infection by this bacterium. Although Ishida *et al.* (26) reported stable maize transformants following infection by *A. tumefaciens*, most successful transformations of maize plants have used microprojectile bombardment to deliver DNA to embryogenically responsive cells from immature embryos. Since the first report of the generation of stable transgenic maize plants using the biolistic method (27), this technique has been used to introduce a number of agronomically important traits into corn, such as insect resistance (20, 28), viral resistance (29), and fructan production (30). These advances, along with our report, attest to a promising future for the use of transgenic technologies in the genetic study and agronomic improvement of maize.

We thank Hank Bass for helpful discussions at the inception of this work, Susan Martino-Catt for technical assistance, and Becky Boston for providing the BiP antibody. This research was supported by a grant from the Department of Energy to B.A.L. (DEFG0292ER 20079) and a postdoctoral fellowship from the National Institutes of Health to C.E.C.

1. Thompson, G. A. & Larkins, B. A. (1989) *BioEssays* **10,** 108–113.
2. Lending, C. R. & Larkins, B. A. (1989) *Plant Cell* **1,** 1011–1023.
3. Esen, A. & Stetler, D. A. (1992) *Am. J. Bot.* **79,** 243–248.
4. Mertz, E. T., Bates, L. S. & Nelson, O. E. (1964) *Science* **145,** 279–280.
5. Nelson, O. E., Mertz, E. T. & Bates, L. S. (1965) *Science* **150,** 1469–1470.
6. Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. & Thompson, R. D. (1989) *EMBO J.* **8,** 2795–2801.
7. Kodrzycki, R., Boston, R. S. & Larkins, B. A. (1989) *Plant Cell* **1,** 105–114.
8. Schmidt, R. J., Burr, F. A., Aukerman, M. J. & Burr, B. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 46–50.
9. Geetha, K. B., Lending, C. R., Lopes, M. A., Wallace, J. C. & Larkins, B. A. (1991) *Plant Cell* **3,** 1207–1219.
10. Soave, C., Dossena, S., Lorenzoni, C., Di Fonzo, N. & Salamini, F. (1978) *Maydica* **23,** 145–152.
11. Lending, C. R. & Larkins, B. A. (1992) *Protoplasma* **171,** 123–133.
12. Fontes, E. B. P., Shank, B. B., Wrobel, R. L., Moose, S. P., O'Brian, G. R., Wurtzel, E. T. & Boston, R. S. (1991) *Plant Cell* **3,** 483–496.
13. Boston, R. S., Fontes, E. B. P., Shank, B. B. & Wrobel, R. L. (1991) *Plant Cell* **3,** 497–505.
14. Marocco, A., Santucci, A., Cerioli, S., Motto, M., Di Fonzo, N., Thompson, R. D. & Salamini, F. (1991) *Plant Cell* **3,** 507–515.
15. Zhang, F. & Boston, R. S. (1992) *Protoplasma* **171,** 142–152.
16. Lopes, M. A., Coleman, C. E., Kodrzycki, R., Lending, C. R. & Larkins, B. A. (1994) *Mol. Gen. Genet.* **245,** 537–547.
17. Coleman, C. E., Lopes, M. A., Gillikin, J. W., Boston, R. S. & Larkins, B. A. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6828–6831.
18. Gillikin, J. W., Zhang, F., Coleman, C. E., Bass, H. W., Larkins, B. A. & Boston, R. S. (1997) *Plant Physiol.* **14,** 345–352.
19. Dennehy, B. K., Petersen, W. L., Ford-Stantino, C. G., Pajeau, M. & Armstrong, C. L. (1994) *Plant Cell Tissue Organ Cult.* **36,** 1–7.
20. Koziel, M. G., Beland, G. L., Bowman, C., Carozzi, N. B., Crenshaw, R., *et al.*, (1993) *Bio/Technology* **11,** 194–200.
21. Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. & Allard, R. W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 8014–8018.
22. Wallace, J. C., Lopes, M. A., Paiva, E. & Larkins, B. A. (1990) *Plant Physiol.* **92,** 191–196.
23. Lending, C. R., Kriz, A. L., Larkins, B. A. & Bracker, C. E. (1988) *Protoplasma* **143,** 51–62.
24. Knecht, D. A. & Dimond, R. L. (1984) *Anal. Biochem.* **136,** 180–184.
25. McElroy, D. & Brettell, R. I. S. (1994) *Trends Biotech.* **12,** 62–68.
26. Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T. & Kumashiro, T. (1996) *Nat. Biotech.* **14,** 745–750.
27. Gordon-Kamm, W. J., Spencer, T. M., Mangano, M. L., Adams, T. R., Daines, R. J., Start, W. G., O'Brien, J. V., Chambers, S. A., Adams, W. R., Willetts, N. G., Rice, T. B., Mackey, C. J., Krueger, R. W., Kausch, A. P. & Lemaux, P. G. (1990) *Plant Cell* **2,** 603–618.
28. Armstrong, C. L., Parker, G. B., Pershing, J. C., Brown, S. M., Sanders, P. R., *et al.*, (1995) *Crop Sci.* **35,** 550–557.
29. Murry, L. E., Elliott, L. G., Capitant, S. A., West, J. A., Hanson, K. K., Scarafia, L., Johnston, S., DeLuca-Flaherty, C., Nichols, S., Cunana, D., Dietrich, P. S., Mettler, I. J., Dewald, S., Warnick, D. A., Rhodes, C., Sinibladi, R. M. & Brunke, K. J. (1993) *Bio/Technology* **11,** 1559–1564.
30. Caimi, P. G., McCole, L. M., Klein, T. M. & Kerr, P. S. (1996) *Plant Physiol.* **110,** 355–363.

THE JOURNAL OF BIOLOGICAL CHEMISTRY
© 1985 by The American Society of Biological Chemists, Inc.

Vol. 260, No. 30, Issue of December 25, pp. 16451-16459, 1985
Printed in U.S.A.

Nucleotide Sequence Analysis of Zein mRNAs from Maize Endosperm*

(Received for publication, May 10, 1985)

M. David Marks, Judith S. Lindell, and Brian A. Larkins‡

From the Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana 47907

A comparison of the DNA and protein sequences of a group of zein cDNA clones reveals that they share extensive sequence homology and probably originated from a common ancestral gene. A comparison of clones corresponding to M_r 22,000 polypeptides shows they are 92% homologous, while five clones corresponding to the M_r 19,000 zeins vary in homology from 75 to 95%. The clones corresponding to the M_r 22,000 proteins are 60–65% homologous to clones encoding the M_r 19,000 zein proteins. A clone corresponding to the M_r 15,000 zein has little homology to either the M_r 22,000 or 19,000 zeins. Clones corresponding to both the M_r 22,000 and 19,000 zeins have two putative polyadenylation signals. S1 nuclease mapping indicates that the first polyadenylation signal following the stop codon is utilized by the M_r 22,000 sequences, while primarily the second polyadenylation signal is utilized by the M_r 19,000 sequences.

To study the expression of zein genes during maize endosperm development, we constructed and characterized a number of full-length cDNA clones (Marks *et al.*, 1985). On the basis of cross-hybridization studies, these clones were divided into nine distinct groups: one for the M_r 15,000 zein, five for the M_r 19,000 zeins, and three for the M_r 22,000 zeins. No clones for the M_r 10,000 or reduced soluble protein were isolated. However, the clones that were identified appear to represent a large proportion of the zein mRNA sequences in the endosperm mRNA population (Marks *et al.*, 1985).

Sequences of clones related to some of those in the groups we describe have been reported (Heidecker and Messing, 1983; Geraghty *et al.*, 1982; Spena *et al.*, 1982). Previous studies revealed that genes encoding the M_r 19,000 and 22,000 zein groups are homologous and probably have a common ancestral sequence (Marks and Larkins, 1982; Spena *et al.*, 1982), but the M_r 15,000 zein sequence was found to belong to an unrelated gene family.[1] Genes within the M_r 19,000 and 22,000 zein groups have diverged from one another through base substitutions, small insertions/deletions, and relatively large internal duplications (Heidecker and Messing, 1983; Marks and Larkins, 1982). Many of these genes contain two functional polyadenylation signals (Messing *et al.*, 1983), and several of the genes appear to have multiple transcriptional start sites (Langride and Feix, 1983).

This study presents a comprehensive comparison of sequences belonging to eight M_r 19,000 and 22,000 zein groups, and it presents the sequence of a full-length cDNA clone for a M_r 15,000 zein that contains a very long 5′ noncoding region. We have also analyzed the usage of the polyadenylation signals for the M_r 19,000 zein genes.

MATERIALS AND METHODS

DNA restriction endonucleases, Klenow fragment, T4 kinase, and S1 nuclease were purchased from Bethesda Research Laboratories. [γ-^{32}P]ATP, [α-^{32}P]dATP, and [α-^{32}P]dCTP were purchased from New England Nuclear or Amersham Corp. Nitrocellulose was obtained from Schleicher and Schuell. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim.

5′-Terminal End Labeling and DNA Sequencing—Restriction fragments were prepared for sequencing by the radioactive end-labeling method of Maxam and Gilbert (1980) for 5′ ends as described by Pedersen *et al.* (1982). DNA fragments were sequenced by the method of Maxam and Gilbert (1980) with the modification of Krayev *et al.* (1980).

Computer Analysis—Sequences were analyzed using the Map and Gap programs developed by Devereux *et al.* (1984).

S1 Nuclease Mapping—Fragments for S1 nuclease mapping were either 5′ end labeled as described by Maxam and Gilbert (1980) or 3′ end labeled by filling in recessed 3′ ends with Klenow fragment as described by Maniatis *et al.* (1982). End-labeled fragments were cleaved by restriction enzyme digestion and resolved on a 5% polyacrylamide gel. The fragments of interest were excised from the gel and isolated by the crush-soak method of Maxam and Gilbert (1980).

S1 nuclease mapping was performed by the method of Berk and Sharp (1977) as described by Maniatis *et al.* (1982). One μg of endosperm poly(A) RNA plus 200 μg of tRNA were added to 3×10^4 to 1×10^6 cpm of labeled fragment in 20 μl of 40 mM Pipes[2] (pH 6.4), 1 mM EDTA (pH 8.0), 0.4 M NaCl, and 80% formamide. Reactions were heated to 80 °C for 10 min and then incubated overnight at 58 °C for Z15 fragments (% G + C = 55%) and at 52 °C for Z19 fragments (% G + C = 48%). The hybridization reaction was stopped by adding 0.38 ml of cold S1 nuclease buffer (0.28 M NaCl, 0.05 M NaOAc (pH 4.6), 4.5 mM $ZnSO_4$, and 20 μg/ml denatured calf thymus DNA) containing 400 units of S1 nuclease. The S1 digestion reactions were incubated at 37 °C for 30 min after which they were extracted with phenol/chloroform and the DNAs precipitated with ethanol. Samples were dissolved in sequencing dye (80% formamide, 0.5 × TNE (10 mM Tris-HCl, pH 7.4, 15 mM NaCl, and 1 mM EDTA), 0.1% bromphenol blue, and 0.1% xylene cyanol) and loaded onto a 6% polyacrylamide sequencing gel that contained 50% urea.

Northern Blot Analysis—Four μg of poly(A) RNA from membrane-bound polysomes of 18 day-after-pollination endosperms was resolved by electrophoresis on an agarose gel that contained methyl mercury hydroxide (1.4% agarose, 5 mM methyl mercury hydroxide, 50 mM boric acid, 5 mM sodium borate, 10 mM sodium sulfate, and 0.01 mM Na_2EDTA) (Bailey and Davidson, 1976). The gel was soaked in 0.5 M ammonium acetate and then stained with ethidium bromide. Gel slices that contained the upper and lower zein mRNA bands were excised and recast in separate 1.4% agarose gels. The RNA was transferred to nitrocellulose filters (Thomas, 1980), and the nitrocellulose was prehybridized for 8 h at 42 °C in 50% formamide, 5 × SSC (1 × SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.05 M sodium phosphate buffer, pH 6.8, 0.1% bovine serum albumin, 0.1% Ficoll,

* This is Journal Paper 10,319 of the Purdue Agricultural Experiment Station. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed.

[1] K. Pedersen, P. Argos, S. V. L. Nayarana, and B. A. Larkins, manuscript in preparation.

[2] The abbreviation used is: Pipes, 1,4-piperazinediethanesulfonic acid.

0.1% polyvinylpyrrolidone, 200 μg/ml heat-denatured sheared calf thymus DNA. The nitrocellulose was then hybridized overnight at 42 °C in 50% formamide, 5 × SSC, 0.05 M sodium phosphate buffer, pH 6.8, 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 200 μg/ml sheared calf thymus DNA, 5% dextran sulfate, and 7×10^6 cpm of the isolated zein cDNA inserts labeled with ^{32}P by nick translation (Maniatis *et al.*, 1976). Following hybridization the nitrocellulose was washed twice in 1 × SSC containing 0.1% sodium dodecyl sulfate at room temperature and twice in 1 × SSC containing 0.1% sodium dodecyl sulfate at 68 °C. Filters were air dried and autoradiographed.

RESULTS

By cross-hybridization analysis at a stringent criterion (T_m −15 °C), we were able to distinguish nine groups of zein cDNA clones (Marks *et al.*, 1985). The DNA sequences of several members from three of these groups was previously reported (Marks and Larkins, 1982; Pedersen *et al.*, 1982).[1] In this study seven cDNA clones from the remaining six groups were completely sequenced by the method of Maxam and Gilbert (Fig. 1).

Analysis of M_r 22,000 Zein Sequences—Each of the clones, cZ22A-1 and cZ22B-1, encodes a complete zein protein, whereas the third, cZ22C-2, lacks a few coding bases on the 5′ end (Fig. 2). Both cZ22B-1 and cZ22C-2 contain an entire 3′ noncoding region, and cZ22A-1 contains a long 5′ noncoding region. The cZ22B-1 and cZ22C-2 sequences align without gaps; however, when these clones were compared to cZ22A-1, four short gaps were introduced to maximize homology. The largest gap results in a 9-base pair deletion in cZ22A-1 relative to cZ22B-1 and cZ22C-2. Overall, the deletion/insertion events have left cZ22B-1 and cZ22C-2 with three more amino acids than cZ22A-1. The clones differ from one another by 6–7%, and these nucleotide differences are fairly evenly distributed throughout the sequences. The longest stretch of perfect homology occurs in the region of the stop codon and is only 56 nucleotides in length. Forty-four per cent of the base changes have resulted in amino acid substitutions. In general, the physical properties of the substituted amino acids are conserved. For example, seven of the 26 amino acid changes that occurred during the divergence of cZ22A-1 and cZ22B-1 involved exchanges between valine and alanine (Fig. 3). A few substitutions have not been conservative. For example, at position 103 (Fig. 3) alanine is replaced with aspartic acid, and at positions 177 and 227 glutamic acid is replaced with glutamine and alanine, respectively. One other substitution resulted in a tryptophan residue at position 258. This amino acid has been found only in zein sequences closely related to the cZ22C group (Spena *et al.*, 1982; Geraghty *et al.*, 1982).

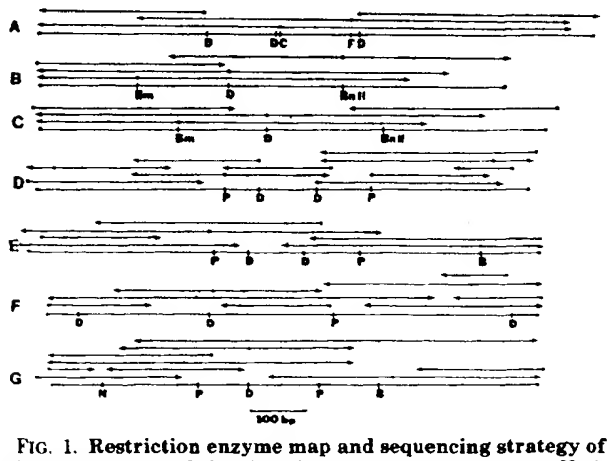


FIG. 1. **Restriction enzyme map and sequencing strategy of zein cDNA clones.** Only selected restriction sites are shown. *Horizontal arrows* indicate the strategy followed for the determination of DNA sequences. The cDNA clones are: *A*, cZ22C2; *B*, cZ19A2; *C*, cZ19B1; *D*, cZ19C1; *E*, cZ19C2; *F*, cZ19D1; and *G*, cZ15A3. Restriction enzymes: *B*, *Ball*; *Bm*, *Bam*HI; *BnII*, *Ban*II; *C*, *Hinc*II; *D*, *Dde*I; *F*, *Hin*fI; *N*, *Nde*I; and *P*, *Pst*I.

Analysis of M_r 19,000 Zein Sequences—We identified five distinct groups of M_r 19,000 zein clones. Members of the four groups Z19A, Z19B, Z19C, and Z19D did not cross-hybridize at a stringent criterion, whereas those of the fifth group, Z19AB, cross-hybridized to members from Z19A and Z19B. The representative member of the Z19AB group, gZ19AB-1, is a subclone of the genomic clone gZ99, whose sequence was previously reported (Pedersen *et al.*, 1982). The remaining groups are represented by cDNA clones, and their sequences are compared to that of gZ19AB-1 (Fig. 4).

Sequences from the groups Z19A, Z19B, and Z19AB are the most closely related and show 94–95% homology. As was the case with the M_r 22,000 zein clones, the nucleotide differences among these sequences show a fairly random distribution (Fig. 4). This is especially true of the cZ19A-2 and cZ19B-1 sequences; the longest stretch of perfect homology between them is less than 50 nucleotides. However, both of these clones have regions of up to 70 nucleotides that are identical to gZ19AB-1. The differences in the distribution of nucleotide changes between clone groups Z19A, Z19B, and Z19AB may explain their cross-hybridization behavior.

To compare members of the same group we determined the DNA sequences of cZ19C-1 and cZ19C-2 (Fig. 4). These two sequences differ by only one nucleotide in the 5′ noncoding region and by four bases over the first 219 coding nucleotides. From that point to the end of the shorter clone, cZ19C-2, the sequences are identical. The four base differences lead to three conservative amino acid differences (Fig. 5).

To align the sequences of Z19A, Z19B, Z19AB, and Z19C groups, several gaps were introduced (Fig. 4). The largest is a 24-base pair deletion in the Z19A, Z19B, and Z19AB sequences (collectively referred to as Z19A/B) relative to the Z19C sequences. The insertion/deletion events that have taken place during the divergence of the genes corresponding to these clones have rendered the Z19C zeins with six more amino acids than the Z19A/B zeins. The Z19C sequences are on the average 85% homologous to the Z19A/B groups. At the 3′ end the average of 85% homology is maintained. However, in the 5′ noncoding region these sequences are only 55–60% homologous.

To align cZ19D-1 with the other sequences, additional gaps had to be introduced (Fig. 4). The longest is a 24-base pair deletion in Z19D-1. This sequence has diverged more than the others and is only 75% homologous.

Five of the six clones encode complete zein polypeptides (Fig. 5). The remaining clone, cZ19A-2, is missing nucleotides for the first three codons. All of the complete polypeptides contain a 21-amino acid signal peptide and an internal tandemly repeated peptide of approximately 20 amino acids that has been previously described (Argos *et al.*, 1982). The polypeptides range in size from 233 to 240 amino acids. As in the case of the M_r 22,000 zein clones, approximately 45% of the nucleotide differences among these clones result in amino acid substitutions. Most of the resulting changes are conservative; however, at a few positions neutral amino acids have been replaced with charged, and vice versa.

Northern Blot Analysis—Previous analyses revealed that the mRNAs encoding the M_r 22,000 zeins can be resolved from the mRNAs for the M_r 15,000 and 19,000 zeins by gel electrophoresis under denaturing conditions (Wienand and Feix, 1978; Larkins *et al.*, 1983). However, the mRNAs for the

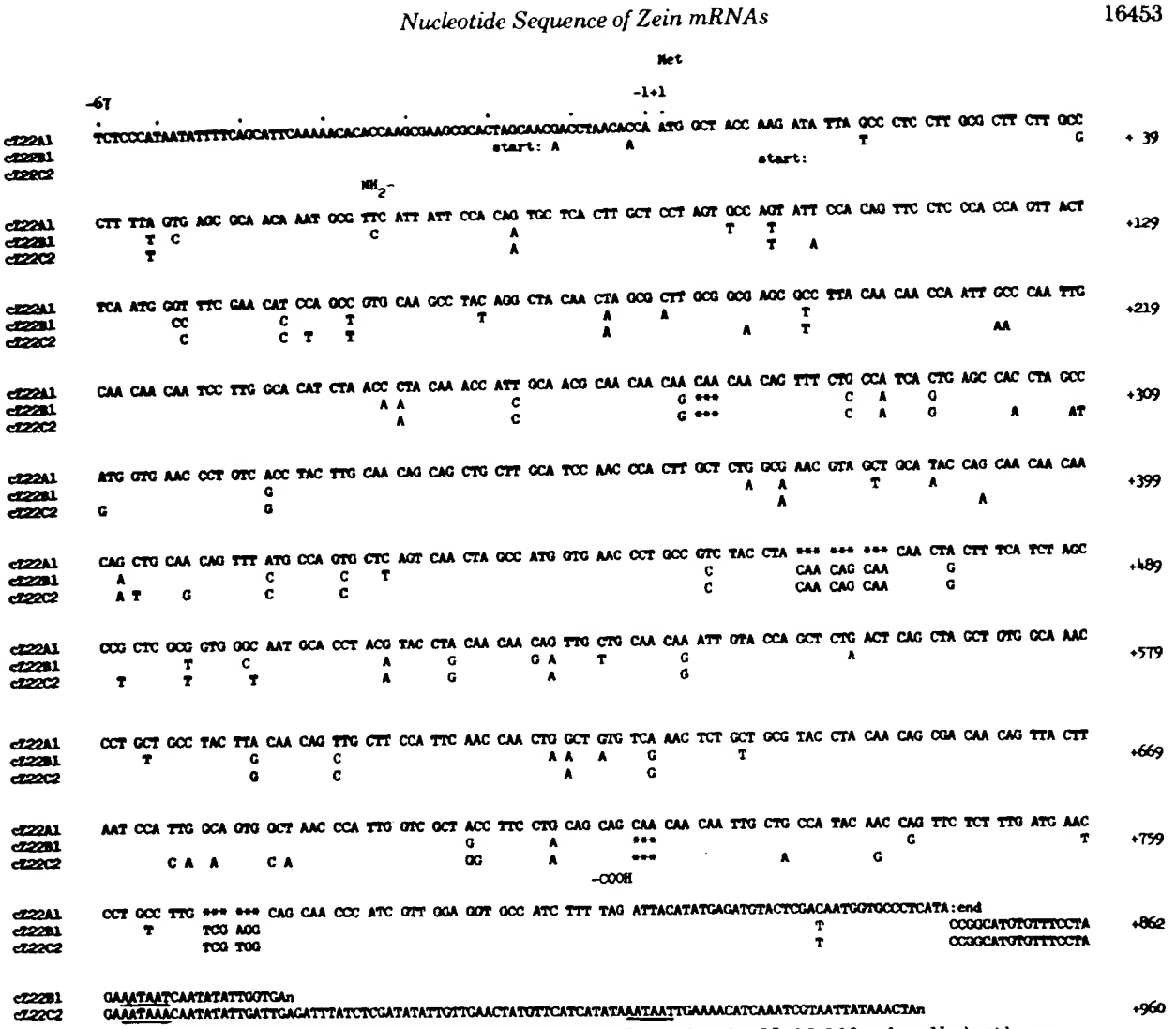


FIG. 2. **Comparison of the nucleotide sequence of clones for the M_r 22,000 zeins.** Nucleotides are numbered starting with the first base of the initiation codon and are indicated on the *right-hand* margin. The sequences of cZ22A1 and cZ22B1 (previously named pZ22.1 and pZ22.3, respectively) were previously reported (Marks and Larkins, 1982). The complete nucleotide sequence of cZ22A1 is given, but only variable nucleotides at corresponding positions are listed for the other clones. *Asterisks* indicate positions where we have introduced gaps in the sequence to maximize homology. The positions of the first nucleotides for cZ22B-1 and cZ22C-2 follow the colons. *An* indicates the position at which a sequence terminates in a poly(A) tail. Positions corresponding to the initiating methionine (Met) and the NH_2 terminus and COOH terminus of the encoded polypeptide are indicated.

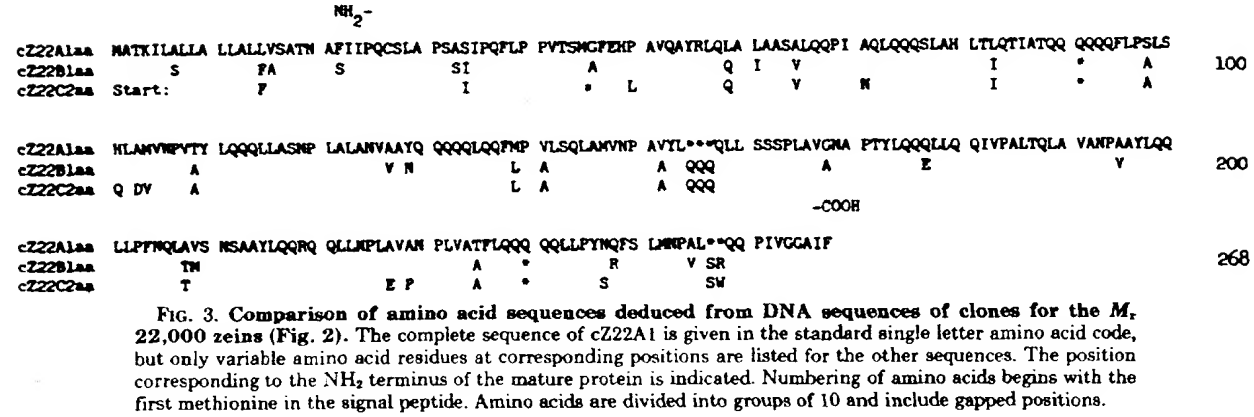


FIG. 3. **Comparison of amino acid sequences deduced from DNA sequences of clones for the M_r 22,000 zeins (Fig. 2).** The complete sequence of cZ22A1 is given in the standard single letter amino acid code, but only variable amino acid residues at corresponding positions are listed for the other sequences. The position corresponding to the NH_2 terminus of the mature protein is indicated. Numbering of amino acids begins with the first methionine in the signal peptide. Amino acids are divided into groups of 10 and include gapped positions.

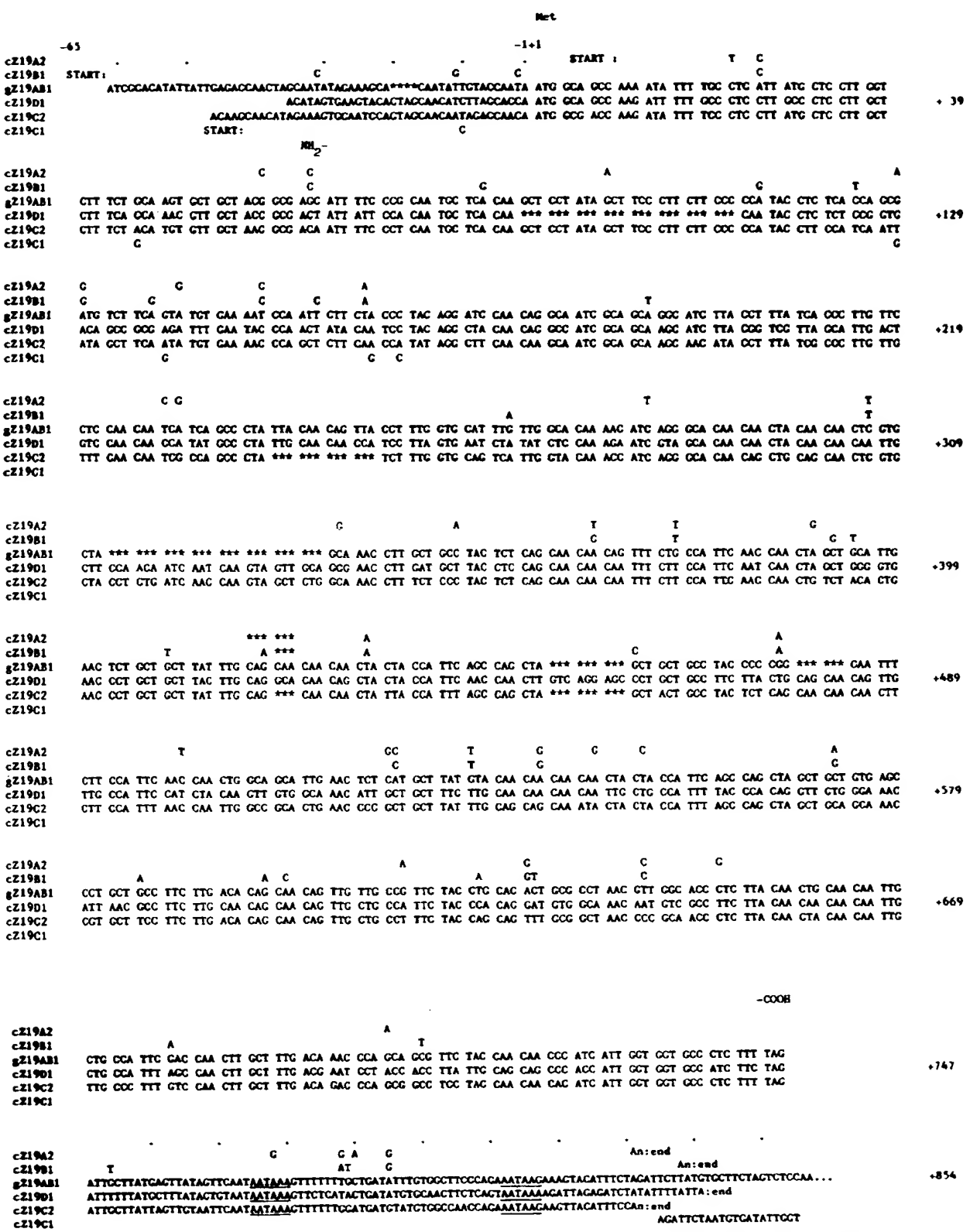


FIG. 4. **Comparison of the nucleotide sequence of clones for the M_r 19,000 zeins.** The nucleotide sequence of gZ19AB1 (previously named ZG99) was previously reported (Pedersen *et al.*, 1982). The complete sequences of gZ19AB1, cZ19D1, and cZ19C2 are given. The sequences of the cZ19A2 and cZ19B1 follow the *colons*, and only nucleotides in the sequences that vary from gZ19AB1 are shown. Likewise, the cZ19C1 sequence begins after the *colon* and only nucleotides in that sequence which vary from cZ19C2 are shown. *An* indicates a sequence terminating in poly(A) tail. *Asterisks* indicate positions where we have introduced gaps in the sequences to maximize homology. The putative polyadenylation signals are *underlined*. Positions corresponding to the initiating methionine (Met) and the NH_2 terminus and COOH terminus of the encoded polypeptide are indicated.

```
cZ19A2  START :    FL              T            T           V          Q                        P
cZ19B1              L              T                      S V          Q
cZ19AB1     MAAKIFCLIM LLGLSASAAT ASIFPQCSQA PIASLLPPYL SPAMSSVCEN PILLPYRIQQ AIAAGILPLS PLFLQQSSAL LQQLPLVHLL AQNIRAQQLQ
cZ19D1      MAAKIFALLA LLALSANVAT ATIIPQCSQ* *******QYL SPVTAARFEY PTIQSYRLQE AIAASILRSL ALTVQQPYAL LQQPSLVNLY LQRIVAQQLQ    100
cZ19C1      MATKIFSLLM LLALSACVAN ATIFPQCSQA PIASLLPPYL PSMIASVCEN PALQPYRLQQ AIAASNIPLS PLLFQQSPAL ****SLVQSL VQTIRAQQLQ
cZ19C2                               T                              I   I

cZ19A2                   G         H                   * *  Q               Q                 A   L      P       D
cZ19B1                                        GS    S    *  Q          P    Q                 P   L              G    T
cZ19AB1     QLVL****** **ANLAAYSQ QQQFLPFNQL AALNSAAYLQ QQQLLPFSQL ***AAAYPR* *QFLPFNQLA ALNSHAYVQQ QQLLPFSQLA AVSPAAFLTQ
cZ19D1      QQLLPTINQV VAANLDAYLQ QQQFLPFNQL AGVNPAAYLQ AQQLLPFNQL VRSPAAFLLQ QQLLPFHLQV VANIAAFLQQ QQLLPFYPQV VGNINAFLQQ    200
cZ19C1      QLVLPLINQV ALANLSPYSQ QQQFLPFNQL STLNPAAYLQ *QQLLPFSQL ***ATAYSQQ QQLLPFNQLA ALNPAAYLQQ QILLPFSQLA AANRASFLTQ
cZ19C2

cZ19A2               A    A  V           N        T
cZ19B1      P      Q V    A              N         V
cZ19AB1     QQLLPFYLHT APNVGTLLQL QQLLPFDQLA LTNPAAFYQQ PIIGGALF
cZ19D1      QQLLPFYPQD VANNVAFLQQ QQLLPFSQLA LTNPTTLLQQ PTIGGAIF                                                     248
cZ19C1      QQLLPFYQQF AANPATLLQL QQLLPFVQLA LTDPAASYQQ HIIGGALF
cZ19C2
```

FIG. 5. **Comparison of amino acid sequences deduced from the DNA sequence of clones for the M_r 19,000 zeins (Fig. 4).** The complete amino acid sequences deduced from gZ19AB1, cZ19D1, and cZ19C1 are given. Only amino acid residues for cZ19A2 and cZ19B1 that differ from gZ19AB1 are shown as are only amino acid residues for cZ19C2 that differ from the sequence of cZ19C1. *Asterisks* indicate gaps in the amino acid sequence. The positions of the NH_2 terminus and COOH terminus of the mature proteins are indicated. The number of amino acids begins with the first methionine in the signal peptide. Amino acids are divided into groups of 10 and include gapped positions.

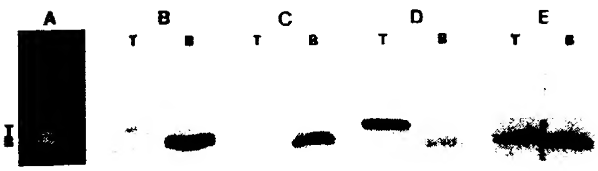


FIG. 6. **Northern blot analysis of zein mRNAs.** *A*, 18 DAP endosperm poly(A) RNA from membrane-bound polyribosomes was resolved on 1.4% agarose gels that contained 5 mM methyl mercury. RNA from gel slices that contained the top (*T*) and bottom (*B*) zein mRNA bands was transferred to nitrocellulose and hybridized to nick-translated zein cDNA inserts as described under "Materials and Methods." The cDNA probes were: *B*, cZ15A3; *C*, cZ19C1; *D*, cZ22B1; and *E*, gZ19AB1.

M_r 15,000 and 19,000 zeins have not been resolved. We used a modified Northern blot analysis to determine if our clone groups correspond to distinct mRNA populations encoding the M_r 22,000, 15,000, or 19,000 zeins.

Poly(A) rRNA was isolated from membrane-bound polysomes of maize endosperm and resolved on a methyl mercury agarose gel (Fig. 6*A*). The upper and lower bands corresponding to zein mRNAs were excised, blotted onto separate nitrocellulose filters, and hybridized to representative clones. Probes representing the Z15 and Z19C groups hybridized only to the lower band (Fig. 6, *B* and *C*). The cZ22B-1 probe hybridized only to the upper band under conditions that allowed cross-hybridization between the three closely related M_r 22,000 zein groups (Fig. 6*D*). These results indicate that the mRNA populations corresponding to the respective clone groups are homogenous and probably encode only a single size class of protein. The 19AB probe hybridized to both the upper and lower bands (Fig. 6*E*), a result which is consistent with the existence of two subsets of mRNAs corresponding to this group (Hu *et al.*, 1982; Heidecker and Messing, 1983). Presumably, one group of mRNAs encodes M_r 19,000 zeins and the other encodes M_r 22,000 zeins.

Analysis of a M_r 15,000 Zein cDNA Clone—We recently compared the sequence of a cDNA clone and a genomic clone encoding a M_r 15,000 zein.[1] The cDNA clone lacks a portion of the 5′ coding region but contains the entire 3′ noncoding sequence of its template mRNA. Except for one nucleotide difference, the cDNA sequence is identical to the genomic sequence for its entire length.

We report here the sequence of an additional cDNA clone that encodes a complete M_r 15,000 zein polypeptide (Fig. 7). This sequence differs from the genomic clone[1] by a few nucleotides. Since we estimated that there may only be two genes for the M_r 15,000 zeins (Wilson and Larkins, 1984), these two sequences may represent them. However, some of the base differences between the clones may be due to sequencing artifacts. Because of the high G + C content, many regions of the sequencing gels were compressed, making the sequence determination extremely difficult. Essentially all of the base differences were in these regions.

Previous S1-mapping data suggested that the 5′ end of the mRNA for the M_r 15,000 gene began 65–70 nucleotides before the start codon (see Fig. 7). However, this larger cDNA clone has over 150 noncoding nucleotides at the 5′ end that align perfectly with the 5′ noncoding sequence of the genomic clone. The existence of this cDNA clone indicates that transcription may originate from two regions of the gene. To investigate this, we performed an S1-mapping analysis. A *Hin*dIII-*Dde*I fragment that contains the 5′ end of the mRNA sequence was hybridized to poly(A) RNA isolated from staged endosperms (Fig. 8*A*) and the hybrids digested with S1 nuclease. The majority of protected fragments ended in the same region previously shown to be the 5′ end of the mRNA. However, upon longer exposure of the gel we detected some transcripts that protected the complete noncoding region of the longer cDNA clone (Fig. 8*B*). While these results are consistent with the possibility of a major and minor origin of transcription, they do not exclude processing of a larger transcript for this gene. The longer transcript appears to be more abundant at early stages of endosperm development, as is indicated by the greater protection of the larger fragment with poly(A) RNA from 12 DAP endosperms than from later stages (Fig. 8*B*). Further analysis of the transcriptional activity of the M_r 15,000 gene will be presented elsewhere.[3]

The last 23 nucleotides of the 3′ end of the cZ15A3 were not homologous to the genomic clone. The terminal 28 nucleotides were found to be an inverted copy of the sequence

[3] R. S. Boston and B. A. Larkins, manuscript in preparation.

```
                                                                                                    XXX
cZ15A3  AATCCATGTACTATGTTTCCACGTCATGCAACGCAACATTCCAAAACCATGGATCATCTATAAATGGCTAGCTCCCACATATGAACTAGTCTCTATCATCATCCAATCCAGATCAGCAAA   - 39

                                                -1 +1
cZ15A3  GCGGCAGTGCGTAGAGAGGATCGTCGAACCAGAACAGC  ATG AAG ATG GTC ATC GTT CTC GTC GTG TGG CTG GCT CTG TCA GCT GCC AGC GCC TCT GCA   + 60
                                                 M   K   M   V   I   V   L   V   V   W   L   A   L   S   A   A   S   A   S   A

        NH2-
cZ15A3  ATG CAG ATG CCC TGC CCC TGC GCG GGG CTG CAG GGC TTG TAC GGC GCT GGC GCC GGC CTG ACG ACG ATG ATG GGC GCC GGC GGG CTG TAC   +150
         M   Q   M   P   C   P   C   A   G   L   Q   G   L   Y   G   A   G   A   G   L   T   T   M   M   G   A   G   G   L   Y

cZ15A3  CCC TAC GCG GAG TAC CTG AGG CAG CCG CAG TGC AGC CCG CTG GCG GCG GCG CCC TAC TAC GCC GGG TGT GGG CAG ACG AGC GCC ATG TAC   +240
         P   Y   A   E   Y   L   R   Q   P   Q   C   S   P   L   A   A   A   P   Y   Y   A   G   C   G   Q   T   S   A   M   Y

cZ15A3  CAG CCG CTC CGG CAA CAG TGC TGC CAG CAG CAG ATG AGG ATG ATG GAC GTG CAG TCC GTC GCG CAG CAG CTG CAG ATG ATG ATG CAG CTT   +330
         Q   P   L   R   Q   Q   C   C   Q   Q   Q   M   R   M   M   D   V   Q   S   V   A   Q   Q   L   Q   M   M   M   Q   L

cZ15A3  GAG CGT GCC GCT ACC GCC AGC AGC AGC CTC TAC GAG CCA GCT CTG ATG CAG CAG CAG CAG CAG CTG CTG GCA GCC CAA GGT CTC AAC CCC   +420
         E   R   A   A   T   A   S   S   S   L   Y   E   P   A   L   M   Q   Q   Q   Q   Q   L   L   A   A   Q   G   L   N   P

cZ15A3  ATG GCC ATG ATG ATG GCG CAG AAC ATG CCG GCC ATG GGT GGA CTC TAC CAG TAC CAG TAC CAG CTG CCC AGC TAC CGC ACC AAC CCC TGT   +510
         M   A   M   M   M   A   Q   N   M   P   A   M   G   G   L   Y   Q   Y   Q   Y   Q   L   P   S   Y   R   T   N   P   C

                                                -COOH
cZ15A3  GGG GTC TCC GCT GCC ATT CCG CCC TAC TAC TGA  TTCATGATATTTGGGAAATCTCCTCTATCCATCCCTCTCTATCTATATATGTAATAATGCAGTAAGACGACACA   +617
         G   V   S   A   A   I   P   P   Y   Y   •

cZ15A3  CATTATCATGTGTGGTATGACCAATAATATATGCATCATAATAAAGTTTTGGTCATACCACACATGATAATGTGTG
```

FIG. 7. **Nucleotide sequence of a cDNA clone encoding a M_r 15,000 zein protein.** The predicted amino acid sequence is shown below the nucleotide sequence. *XXX* indicates the major transcriptional start site, and the *underscored* sequence corresponds to the putative polyadenylation signal. The *overscored* region indicates the position of an inverted nontandem repeated sequence. The NH_2-terminal and COOH-terminal positions of the mature polypeptide are indicated.

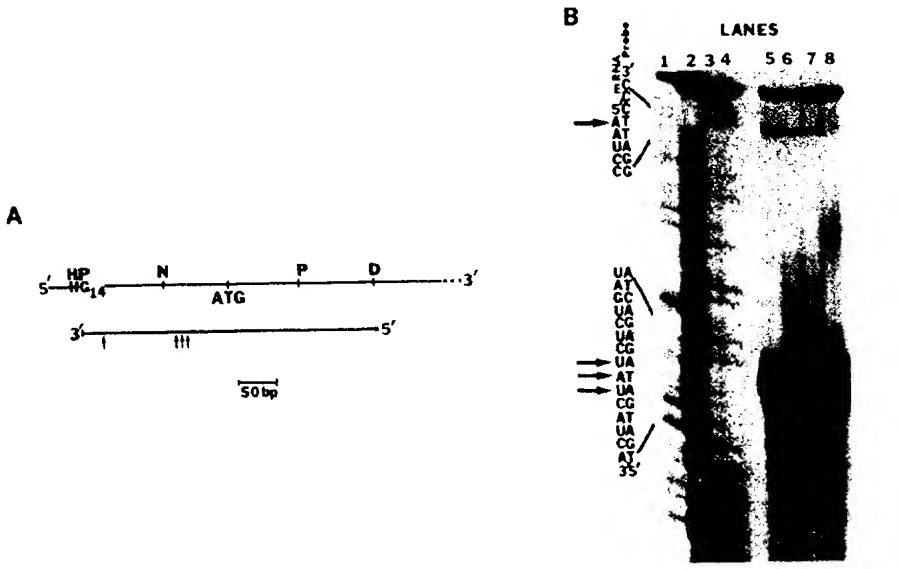


FIG. 8. **S1 mapping of the 5′ region of cZ15A3.** *A*, a partial restriction enzyme map of the 5′ end of cZ15A3 and a region of the clone used in the S1-mapping experiment. G_{14} indicates a G tract of 14 bases that was added during the cloning procedure. The restriction enzyme sites upstream from the G tract are in the poly linker region of the plasmid vector pUC8. The probe was generated by end labeling a *Dde*I fragment with subsequent cleavage at the *Hin*dIII site in the poly linker. The *vertical arrows* indicate the limits of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in *lane 1* (G reaction), *lane 2* (G + A reaction), *lane 3* (C + T reaction), and *lane 4* (C reaction). The probe was hybridized to 1 μg of poly(A) RNA from 12 DAP (*lane 5*), 18 DAP (*lane 6*), 22 DAP (*lane 7*), and 28 DAP (*lane 8*) endosperms as described under "Materials and Methods." After the hybridization the probe was digested with S1 nuclease and resolved on a sequencing gel. The sequence in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given on the *left-hand* margin. Restriction enzyme key: *D*, *Dde*I; *H*, *Hin*dIII; *N*, *Nde* I; and *P*, *Pst*I.

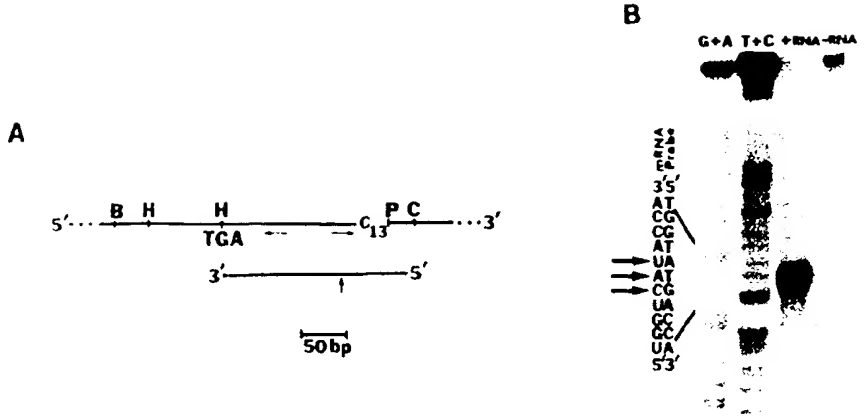


FIG. 9. **Mapping of the 3′ region of cZ15A3.** *A*, a partial restriction enzyme map of the 3′-terminal of cZ15A3 and the region of the clone used in the S1-mapping experiment. The C_{13} represents a C tract of 13 bases that was added during the cloning procedure (Marks *et al.*, 1985). The restriction enzyme sites downstream from the C tract are in the poly linker region of the plasmid vector, pUC8. The probe was generated by cleavage with *Hin*fI, and it was 3′ end labeled by using Klenow fragment. The labeled fragment was recut with *Hin*cII. The *horizontal arrows* show the position of inverted repeats (see Fig. 6), and the *vertical arrows* indicate the limit of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in the *first two lanes* as indicated. The probe was hybridized to 1 μg of 18 DAP endosperm poly(A) RNA plus 200 μg tRNA (+RNA) or 200 μg of tRNA without poly(A) RNA (−RNA) as described under "Materials and Methods." After hybridization the probe was digested with S1 nuclease and resolved on a sequencing gel. The sequence in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given at the *left-hand* margin. Restriction enzymes: *B*, *Bal*I; *C*, *Hin*cII; *H*, *Hin*fI; and *P*, *Pst*I.

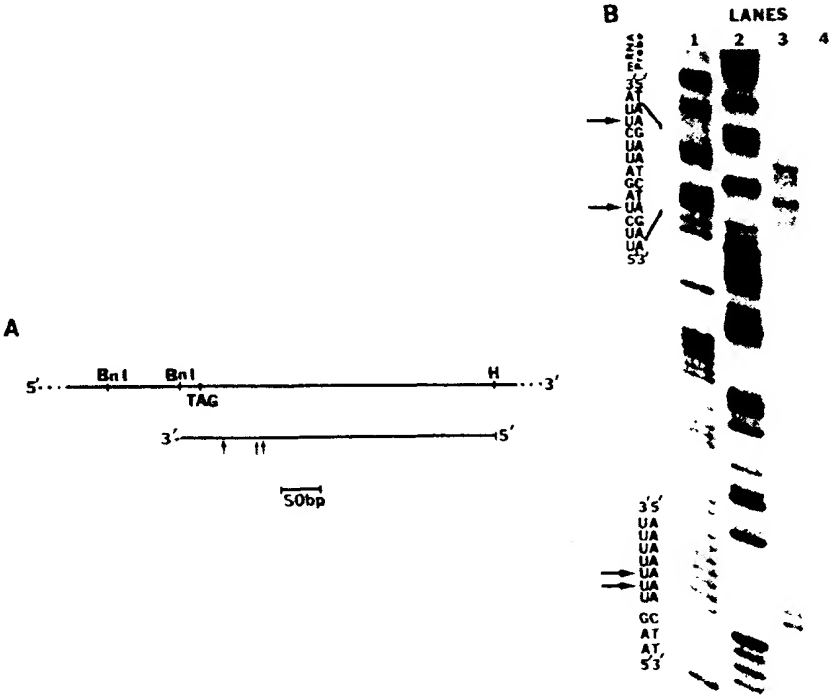


FIG. 10. **Determination of the 3′ terminus of mRNAs closely related to the Z19A/B clones.** *A*, a partial restriction enzyme map of the 3′ coding and noncoding sequence of the clone gZ19AB-1 (also called ZG99, Pedersen *et al.*, 1982) and the region of the clone used in a S1-mapping experiment. The probe was made by cleavage with *Ban*I followed by 3′ end labeling with Klenow fragment and a second restriction digestion with *Hin*cII. The *vertical arrows* indicate the limits of S1 digestion as shown in *B*. *B*, Maxam and Gilbert sequencing reactions of the probe are shown in *lane 1* (G + A reaction) and *lane 2* (C + T reaction). The probe was hybridized to 1 μg of 18 DAP endosperm poly(A) RNA plus 200 μg of tRNA (*lane 3*) and to 200 μg of tRNA alone (*lane 4*) as described under "Materials and Methods." After hybridization the probe was digested with S1 nuclease and resolved on a squencing gel. The sequences in the region of the nuclease-protected bands and the location of the bands (*arrows*) are given on the *left-hand* margin of *B*. Restriction enzymes: *Bn*I, *Ban*I and *H*, *Hin*cII.

found between nucleotides 614 and 642 (Fig. 7). To determine if the 3′ end of the cDNA is authentic, we used a *Hin*fI to *Hin*cII fragment 3′ end labeled at the *Hin*fI site to probe endosperm poly(A) RNA (Fig. 9*A*). The only protected fragments extended a few bases into the terminal inverted repeat (Fig. 9*B*). Thus, this inverted repeat is apparently a cloning artifact and is not due to sequence heterogeneity.

Analysis of Polyadenylation—Many of the M_r 19,000 zein genes were found to contain two polyadenylation signals (Messing *et al.*, 1983). To quantitate the relative utilization of these signals in the genes belonging to the Z19A/B groups (Fig. 4) we did an S1-mapping experiment. A *Ban*I to *Hin*cII fragment from the genomic clone 19AB-1 was 3′ end labeled on the *Ban*I site (Fig. 10*A*) and hybridized to endosperm poly(A) RNA (Fig. 10*B*). The probe should be protected by mRNA corresponding to the Z19A/B groups since these sequences differ by only a few nucleotides in this region. Two major fragments that differ in size by seven nucleotides were protected (Fig. 10*B*). Both of these fragments extend past the second polyadenylation signal. Furthermore, the shorter and

longer fragments end precisely at the last nucleotides found in cZ19A-2 and cZ19B-1, respectively (Fig. 4). Several additional bands, fainter in intensity, correspond to fragments ending three bases after the first polyadenylation signal. Since addition of the poly(A) tail normally occurs within 10–30 nucleotides of the polyadenylation sequence (Nevins, 1983), these may not represent authentic 3′ ends of the transcripts. They may instead represent a region of instability due to a long tract of Ts in this area. A few faint bands corresponding to mRNAs terminating 10–40 nucleotides beyond the first polyadenylation signal also appeared.

Heidecker and Messing (1983) have sequenced two cDNAs closely related to 19A/B groups that terminate after the first polyadenylation signal. Thus, the first site is probably used, but at a much lower frequency than the second. Similar results were obtained with probes from the 19C and 19D groups (data not shown).

DISCUSSION

To study the heterogeneity among the zein mRNAs in the inbred W64A, we characterized the structure of a number of zein cDNA clones. Each of the three groups of M_r 22,000 zein sequences is 60–70% homologous to the M_r 19,000 zein sequences. Sequences among the M_r 22,000 groups showed 92% homology, and sequences among the M_r 19,000 groups have homology ranging from 75 to 95%. The relationships among all these zein sequences are illustrated in Fig. 11, which suggests the order of duplications that gave rise to these groups.

Hu *et al.* (1982) characterized a genomic clone, Z4, closely related to our Z19A/B groups that contains a large internal duplication. The duplication adds 30 amino acids to the encoded polypeptide and makes it the size of a M_r 22,000 zein. By Northern blot analysis we found that in the W64A inbred this type of duplication appears to be restricted to genes within the Z19A/B group (Fig. 6).

The duplication event suggests that the M_r 22,000 zeins may have originated from a M_r 19,000 gene, and close examination of the M_r 22,000 sequences provides evidence for this. The sequence from nucleotides 264–396 is 85% homologous to the sequence from nucleotides 397–540. The duplication that formed this region would have added approximately 40 amino acids to the encoded polypeptide. A series of base substitutions would account for the subsequent diversion of the genes into two distinct gene families.

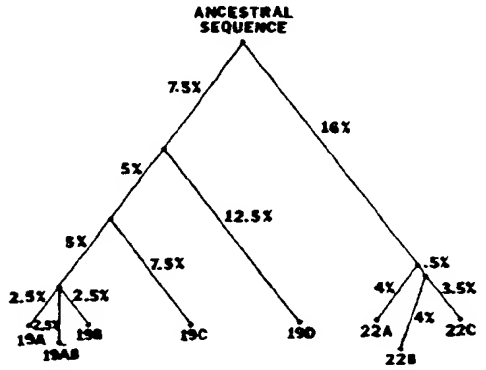


FIG. 11. **Comparison of sequence homology among zein mRNAs.** The phylogeny was reconstructed with the method described by Fitch and Margoliash (1967) from the approximate per cent of nucleotide substitutions that have taken place among the zein genes. The percentages listed are the approximate per cent of nucleotide substitution between putative gene branching points.

We have also sequenced a cDNA clone for the M_r 15,000 zein which is nearly identical to that obtained for a genomic clone encoding this protein.[1] The sequences of the M_r 15,000 zeins show no homology to those for either the M_r 19,000 or 22,000 zeins, indicating that they comprise a separate family of genes. The M_r 15,000 zeins have a much higher G + C content, lack an internal repeating nucleotide sequence, and are present in only one or two copies in the genome (Wilson and Larkins, 1984). According to S1 mapping, the mRNA encoding the M_r 15,000 zein primarily begins 65–70 nucleotides before the start codon (Fig. 8). The sequence of cZ15A3 differs from this by having a 5′ noncoding sequence of over 150 nucleotides. Based on S1 mapping we estimate that the longer mRNA is 100–1000-fold less abundant than the major mRNA species (Fig. 8). Typical "TATA" boxes (Efstratiadis *et al.*, 1980) are found in the genomic clone just upstream from the start of both mRNA species suggesting that both start sites may be used. But it is also possible that the larger transcript is a precursor of the smaller one. The existence of zein genes with two transcriptional promoters is not without precedence. Langride and Feix (1983) analyzed a gene that encodes a M_r 22,000 zein that appears to have two promoters. However, these promoters are separated by over 1000 nucleotides, and they have nearly equal levels of transcriptional activity.

Of the five zein cDNA clones that contained poly(A) tails only three had the consensus sequence "AATAAA" which is found on the 3′ terminus of many animal genes (Nevins, 1983). This sequence was located 50–100 nucleotides before the start of the poly(A) tail, whereas in animal genes the polyadenylation signal is generally 10–30 nucleotides upstream from the poly(A) tail. Messing *et al.* (1983) identified several putative variants of the polyadenylation signal in zein cDNA sequences that were located closer to the poly(A) tail. In the A20 and A30 clones, which are closely related to our Z19C and C19A/B groups, respectively, they found the sequence "AATAAG." We also found this sequence on the 3′ terminus of the Z19C and Z19A/B clones (Fig. 4). In their B49 sequence (closely related to the Z22C group) another putative variant, "AATAAT," was found (Geraghty *et al.*, 1982). This sequence is also found on the 3′ terminus of our Z22B and Z22C clones (Fig. 2).

In addition to the variation in the sequence of the polyadenylation signal, some plant genes also differ from animal genes by having two potential polyadenylation sites (Messing *et al.*, 1983). Hu *et al.* (1982) reported the sequence of a genomic clone, Z4, that has a "AATAAA" sequence and a "AATAAG" sequence 26 and 65 bases, respectively, downstream from the stop codon. They also isolated cDNA clones that terminated in a poly(A) tail after each of these sequences (Heidecker and Messing, 1983). These clones are closely related to our cZ19A2 and cZ19B1 clones which end in a poly(A) tail after the second putative polyadenylation signal.

Our results indicated that most of the zein mRNA that corresponds to the Z19A/B clones terminated after the second polyadenylation sequence. This appears to be in contrast to mRNAs closely related to the M_r 22,000 clone groups (Z22A, B, C). Three of four of the cDNA clones which belong to these groups represented mRNAs that were polyadenylated after the first polyadenylation sequence (Geraghty *et al.*, 1982; Spena *et al.*, 1982). The remaining clone, cZ22C-2, represented an mRNA that was polyadenylated after a second polyadenylation site (Fig. 2). Interestingly, the sequence that Spena *et al.* (1982) determined for the cDNA clone, za1, is identical to that of cZ22C2 up to the point at which za1 ends in a poly(A)

tail. Thus, it appears as though there are preferred polyadenylation sites, but their utilization is variable.

REFERENCES

Argos, P., Pedersen, K., Marks, M. D., and Larkins, B. A. (1982) *J. Biol. Chem.* **257,** 9984–9990

Bailey, J. M., and Davidson, N. (1976) *Anal. Biochem.* **70,** 75–85

Berk, A. J., and Sharp, P. A. (1977) *Cell* **12,** 721–732

Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slighton, J. L., Blechl, A. E., Smithies, O., Barrelle, E. E., Shoulders, C. C., and Proudfoot, N. J. (1980) *Cell* **21,** 653–668

Fitch, W. M., and Margoliash, E. (1967) *Science* **155,** 279–284

Geraghty, D. E., Messing, J., and Rubenstein, I. (1982) *EMBO J.* **1,** 1329–1335

Heidecker, G., and Messing, J. (1983) *Nucleic Acids Res.* **11,** 4891–4906

Hu, N.-T., Peifer, M. A., Heidecker, G., Messing, J., and Rubenstein, I. (1982) *EMBO J.* **1,** 1337–1342

Krayev, A. S., Kramerov, D. S., Skryabin, K. G., Ryskov, A. P., Bayer, A. A., and Georgiev, G. P. (1980) *Nucleic Acids Res.* **8,** 1201–1215

Langride, P., and Feix, G. (1983) *Cell* **34,** 1015–1022

Larkins, B. A., Pedersen, K., Marks, M. D., Wilson, D. R., and Argos, P. (1983) in *Structure and Function of Plant Genomes* (Ciferri, O., and Dure, L., III, eds) pp. 78–83, Plenum Publishing Corp., New York

Maniatis, T., Hardison, R. G., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., and Efstratiadis, A. (1976) *Cell* **15,** 687–701

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual,* pp. 207–209, Cold Spring Harbor Laboratory, New York

Marks, M. D., and Larkins, B. A. (1982) *J. Biol. Chem.* **257,** 9976–9983

Marks, M. D., Lindell, J. S., and Larkins, B. A. (1985) *J. Biol. Chem.* **260,** 16445–16450

Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560

Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J., and Rubestein, I. (1983) in *Genetic Engineering of Plants* (Hollaender, A., Kosuge, T., and Meredith, C. P., eds) pp. 211–228, Plenum Publishing Corp., New York

Nevins, J. R. (1983) *Annu. Rev. Biochem.* **52,** 441–466

Pedersen, K., Devereux, J., Wilson, D. R., Sheldon, E., and Larkins, B. A. (1982) *Cell* **29,** 1015–1026

Spena, A., Viotti, A., and Pirrotta, V. (1982) *EMBO J.* **1,** 1589–1594

Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 5201–5205

Weinand, U., and Feix, G. (1978) *Eur. J. Biochem.* **92,** 605–611

Wilson, D. R., and Larkins, B. A. (1984) *J. Mol. Evol.* **20,** 330–340